

能動学習の基礎と材料工学への適用例

日野 英逸* 田口 優介**
上野 哲朗*** 小野 寛太****

1. はじめに

材料の科学的性質を研究し、その工学的な応用及び所望の性質を有する材料の開発を目的とする材料科学は、材料の計測及び試作の繰り返しを伴う。この観点からすると、材料科学の研究プロセスはモデルの構築と評価を繰り返すデータ解析や機械学習と共通点があるといえよう。機械学習の代表的な問題設定である教師付き学習は、予め大量の入力データと出力データの組が与えられた状況で入出力の対応関係を近似する問題である。モデルの構築に利用できる入出力データの数があれば増えるほどより正確な入出力関係の近似が可能となるが、入力に対して適切な出力を得るために大きなコストがかかることがある。例えば、農作物を育成する環境(各月の平均気温、投与する肥料の種類と量、気象状況等)を「入力」として、特定の農作物の性質(糖度等)を「出力」とすると、ある特定の入力に対して出力を得るためには数ヶ月から数年単位での時間がかかるであろう。また、後述するX線スペクトル計測実験を行う放射光実験施設では研究者が設備を利用できる時間は非常に限られている。入力に対応する実験設定や計測設定として「このような計測を行う」という計画を立てることは可能であるが、対応する出力を得るためのコストが高く、可能な限り少数の計測で最大限の情報を引き出す工夫が必要である。実験を実施する前に説明変数の種類と説明変数の値、そして必要な計測の回数を入念に設計する技術として、実験計画法と呼ばれる方法論がある⁽¹⁾。一方、既にある程度の量のデータ(入出力の組)を観測しており、それを用いて予測モデルが作られているとき、次にどの入力に対して出力の値を取得して、そのデータをモデルの構築に用いるとよりよい予測が出来るようになるかを議論し、自動的に「次の一手」を打つための方法論が能動学習である。能動学習によって予測器を学習するために利用する例題(サンプル)を適切に選択することで、未知の説明変数に対する

応答変数の予測を誤る確率(汎化誤差)が低減することが理論的にも知られている⁽²⁾。なお、能動学習と非常によく似た方法論としてベイズ最適化がある。ベイズ最適化では、最大化あるいは最小化したい評価関数があり、その評価関数を最大にするような入力を可能な限り少ない試行で見つけることを目的としている。一方、能動学習では最小限のサンプル(入出力の組)を用いて入出力関係を学習することを目的とする。本稿では、まず能動学習の問題設定と代表的なアプローチを簡単に説明する。次に、具体的な材料科学への適用例として、X線磁気円二色性スペクトル測定による物質の磁気モーメントの定量解析を能動学習的な手法によって効率化した例を紹介する。

2. 記法の準備

以下では、統計学において入力から出力を予測するといった問題を考える際に標準的に用いられる回帰分析の用語を採用し、入力を説明変数 x 、出力を応答変数 y と呼び、 x から y を予測する予測器 $f(x)$ を学習する問題を考えることにする。 x, y が取りうる値の範囲をそれぞれ X, Y とする。 Y が離散個の要素を持つ場合には、 x から y を予測する問題は判別問題と呼ばれ、 Y が連続的な実数値を取る場合にはこの問題は回帰問題と呼ばれる。予測器によって説明変数 x に対応して予測された応答変数の予測値を $\hat{y}=f(x)$ と表す。予測器に付随して、説明変数 x を固定したときに応答変数 y が取りうる値の確率(事後確率と呼ぶ) $\Pr(y|x)$ あるいは事後確率密度 $p(y|x)$ も得られるものとする。これは予測をするときに応答変数 y の値を一つ確定的に出力するのではなく、取りうる様々な y の値に「確信度」を与えることに相当し、機械学習においてこうした出力が可能な予測器は良く用いられる。確率変数を大文字 X で表し、その実現値を小文字 x で表すものとする。確率関数を $\Pr(X=x)$ のように確率変数 X が値 x をとる確率を与える関数として定義し、確率密度関数を $p(x)$

* 統計数理研究所; 准教授(〒190-8562 東京都立川市緑町10-3)

** 筑波大学; 大学院博士前期課程 *** 量子科学技術研究開発機構; 主任研究員

**** 高エネルギー加速器研究機構; 准教授

Introduction to Active Learning and Application to Material Engineering; Hideitsu Hino*, Yusuke Taguchi**, Tetsuro Ueno*** and Kanta Ono**** (*The Institute of Statistical Mathematics, Tokyo. **University of Tsukuba, Tsukuba. ***National Institutes for Quantum and Radiological Science and Technology, Hyogo. ****High Energy Accelerator Research Organization, Tsukuba)

Keywords: active learning, Gaussian process, material informatics, classification, regression, machine learning

2018年8月20日受理[doi:10.2320/materia.58.7]

で表す。代表的な確率密度関数として、平均ベクトル μ 、分散共分散行列 Σ を持つ多変量 (d 次元) ガウス分布の確率密度関数は

$$p(x; \mu, \Sigma) = (2\pi)^{-d/2} \sqrt{|\Sigma^{-1}|} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

で定義される。ここで正方行列 A に対して $|A|$ でその行列式を表すものとする。予測器あるいは事後確率は、パラメタ θ を有しており、このパラメタをうまく調整することで「学習」を行うものとする。パラメタを明示するために $f_\theta(x)$ 、 $\Pr_\theta(y|x)$ あるいは $f(x; \theta)$ 、 $\Pr(y|x; \theta)$ のように書くことがある。多くの場合、予測器のパラメタは n 個の学習データ $D = \{(x_i, y_i)\}_{i=1}^n$ に関して、予測の損失関数と呼ばれる関数 $l_\theta(x, y)$ の総和 $l_\theta(D) = \sum_{i=1}^n l_\theta(x_i, y_i)$ を最小にすることで行われる。例えば損失関数としては二乗損失 $(y-f(x))^2$ がよく利用され、この場合の最適なパラメタは

$$\hat{\theta} = \operatorname{argmin}_\theta \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$

として得られる。ここで $\operatorname{argmin}_\theta$ とは、その後続く目的関数を最小にするようなパラメタ θ を返す操作である。こうして得られたパラメタで指定される予測器あるいは事後確率を、将来得られる説明変数に対応する応答変数の予測に用いる。なお、特にパラメタを問題としない場合は θ を省略することもある。新たに応答変数の値を計測する対象の説明変数の実現値のことをクエリと呼ぶことにする。このクエリ(問い合わせ)に対して、実際に実験等を行って応答変数の値を確定させることが計測という操作に対応する。クエリに対応する応答変数の値を定めることを、機械学習の用語を採用して「ラベルを付ける」と表現する。

3. クエリ選択の指針

能動学習を行う際、発行するクエリの選択基準にも様々な考え方がある。紙幅の都合から、代表的かつ直感的に理解しやすい3つの考え方を紹介する。全ての能動学習のシナリオには、各データの情報量尺度の評価が含まれる。そのようなクエリ戦略の手法は数多く提案されている。ここで、 x_A^* はアルゴリズム A によって選択された最適なデータとする。以下の分類は、主に(3)に従う。

(1) 不確実性サンプリング

手元に学習した予測モデル $f(x)$ とそれを用いた条件付き確率 $\Pr(\hat{y}|x)$ の計算手段があるとして、最も予測の自信が無いようなサンプルに対してラベル付けを要求することは直観的にも自然であろう。自信のなさ、すなわち不確実性を定量化する方法は色々と考えられるが、代表的な尺度の一つがエントロピー^{†1}である。

$$x_{\text{H}}^* = \operatorname{argmax}_x -\mathbb{E}[\log \Pr(Y|x)]$$

ここで \mathbb{E} は条件付き分布 $\Pr(y|x)$ に関する期待値を取る操作であり、 y が離散個の値を取る場合には $x_{\text{H}}^* = \operatorname{argmax}_x -$

$\sum_{y \in Y} \Pr(y|x) \log \Pr(y|x)$ で、 y が連続値の場合には $x_{\text{H}}^* = \operatorname{argmax}_x - \int_{y \in Y} p(y|x) \log p(y|x) dy$ である。エントロピー基準は、エントロピーが尤度^{†2}の期待値の負であることから、尤度が小さいと期待されるサンプルにラベルをつけることになり、モデルにとって予測に自信がないデータを積極的に学習に取り入れることになる。自信がない分野の勉強を重点的に行うというイメージであり、我々人間が行う「学習」とも共通しているであろう。同じ考え方に基づくより簡便な方法としては、

$$x_{\text{LG}}^* = \operatorname{argmin}_x \Pr(\hat{y}|x)$$

によってクエリを選択する方法がある。ここで $\hat{y} = \operatorname{argmax}_y \Pr_\theta(y|x)$ は現在の予測器で応答変数を予測するという状況で、最も事後確率が高い応答変数値を表す。すなわち、現状の予測器にとって最も確信度が低く、予測が正しいかどうか自信が無いようなサンプル x に対してラベル y を付与してもらう。

(2) 期待モデル変化

不確実性サンプリングでは説明変数を観測したときの応答変数に関する曖昧さに基づきクエリを選択した。別なアイデアとして、ラベル(応答変数値)が得られ、それを学習サンプルに取り入れてモデルを学習した場合に、最も現在のモデルを大きく変えるようなサンプル(説明変数値)をクエリとして選択することも自然であろう。この方針を、期待モデル変化と呼ぶ。これはモデルの変化を定量的に評価できるようなモデルにならば適用可能なアプローチであり、多くの場合モデルの変化を、新しいサンプル (x, y) を加えたときの、学習の目的関数の勾配の絶対値 $\|\nabla l_\theta(D \cup (x, y))\|_2$ で表す。実際には真のラベル y はわからないので、現在得られている事後分布に関する期待値をとって

$$x_{\text{EMC}}^* = \operatorname{argmax}_x \sum_j \Pr(y_j|x) \|\nabla l_\theta(D \cup (x, y_j))\|_2$$

とする。応答変数が連続値の回帰問題においては上式の和を積分に置き換える。このアプローチは、「これまでの考え方を大きく変えるような出来事を積極的に学ぶ」という態度といえる。

(3) 分散最小化

予測器の学習は、通常は何らかの損失関数 $l_\theta(D)$ の最小化で行う。損失関数を最小にするようなサンプルを直接選択するアプローチも考えられるが、実用上難しいことが多い。一方、例えば連続応答変数を扱う回帰問題において一般的に取られる二乗損失最小化基準では、データ (x, y) の分布に関する

†1 (情報)エントロピー
確率分布に付随する不確かさの尺度。

†2 尤度
あるパラメタに従ってデータが生成されるとした場合、観察結果からみてパラメタの値の尤もらしさ(もっともらしさ)を表す数値のこと。

る二乗損失の期待値に関して次のような関係が成り立つことを示すことが出来る：

$$\text{「二乗損失の期待値」} = \text{「誤差の分散」} + \text{「予測のデータへの当てはまりの悪さ」} + \text{「予測器のパラツキ」}$$

ここで、右辺第一項の誤差の分散は予測器に依存しない部分であり、コントロールは出来ない。また、第二項についても、利用する予測器の種類を一度定めたら自動的に決まるものである。最後の予測器のパラツキは数式では $\mathbb{E}_D[(\hat{y} - \mathbb{E}_D[\hat{y}])^2]$ で与えられる予測結果 $\hat{y} = f_\theta(x)$ の分散であり、これは予測器をうまく調整する、あるいは学習サンプルをうまく選択することで減少させることが出来る量である。なお、 $\mathbb{E}_D[\hat{y}]$ は既に得られた学習データ D の分布に関する期待値を意味する。

3. 応 用 例

(1) 格子構造の分類

まず、上述の不確実性サンプリングに基づく方法でランダムサンプリングよりも効率的にクエリを選択できることを見てみよう。無機材料の X 線回折パターン(ピーク位置)を説明変数として、鉄鋼の格子構造が面心立方格子構造(fcc)であるか体心立方格子構造(bcc)であるかを判別する 2 クラス分類問題を考えよう。bcc, fcc のそれぞれの構造に対応した X 線回折パターンを 30889 件 (fcc 24936 件, bcc 5953 件) 準備した。予め 1000 件 (fcc, bcc 各 500 件) をテストデータとして、残りの 30389 件を学習データ候補とした。前処理として主成分分析により説明変数を 3 次元まで削減してから、ロジスティック回帰⁽⁴⁾によって説明変数に対応する応答変数が fcc あるいは bcc のどちらであるかを確率的に予測する予測器を構築して、テストデータを用いてその予測精度を評価する。

実験設定と結果の紹介に入る前に、確率値を出力できる判別器の構成方法の代表例として、ロジスティック回帰を紹介する。なお、「回帰」という名称ではあるが、ロジスティック回帰は判別問題のためのモデルであることに注意されたい。

(2) ロジスティック回帰

2 クラス判別問題を考えるため、応答変数 y は 0 か 1 の 2 値を取るものとする。線形回帰において、説明変数 \mathbf{x} を観測した時の応答変数 y の条件付き分布を $p(y|\mathbf{x}) = p(y - \mathbf{w}^T \mathbf{x} | \mathbf{w})$ としてモデル化したのと同様に、判別問題では応答変数 y の、 \mathbf{x} を観測したときの条件付き確率を $\Pr(y=1|\mathbf{x})$ をモデル化する。簡単のため、 $p = \Pr(y=1|\mathbf{x})$ とおく。この p の関数として、ロジット関数を次式で定義する：

$$\text{logit}(p) = \log \frac{p}{1-p}$$

これは (0, 1) の間の値を定義域とし、実数の全範囲を値域とする関数である。説明変数 \mathbf{x} の値と $\text{logit}(\Pr(y=1|\mathbf{x}))$ の間に線形な関係

$$\text{logit}(\Pr(y=1|\mathbf{x})) = \mathbf{w}^T \mathbf{x}$$

を仮定する。簡単な計算により、

$$\Pr(y=1|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$$

となることを確認できる。上式右辺はロジスティック関数と呼ばれる。観測データ $D = \{(y_i, \mathbf{x}_i)\}_{i=1}^n$ に含まれる個々のサンプル (y_i, \mathbf{x}_i) において、応答変数 y_i が 1 か 0 かのいずれかなので $\Pr(y=0|\mathbf{x}) = 1 - \Pr(y=1|\mathbf{x})$ に注意すると、このモデルにおいて尤度関数は

$$\begin{aligned} l(\mathbf{w}) &= \prod_{(y_i, \mathbf{x}_i) \in D} \Pr(y_i|\mathbf{x}_i) \\ &= \prod_{(y_i, \mathbf{x}_i) \in D} \left(\frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}} \right)^{y_i} \left(\frac{e^{-\mathbf{w}^T \mathbf{x}_i}}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}} \right)^{1-y_i} \end{aligned}$$

となり、対数尤度関数は

$$L(\mathbf{w}) = \sum_{(y_i, \mathbf{x}_i) \in D} \left(y_i \log \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}} + (1 - y_i) \log \frac{e^{-\mathbf{w}^T \mathbf{x}_i}}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}} \right)$$

となる。対数尤度関数 $L(\mathbf{w})$ を係数 \mathbf{w} に関して最大化することで、最尤推定による係数 \mathbf{w} の推定値 $\hat{\mathbf{w}}$ が得られる。通常の線形回帰とは異なり閉じた形の解は存在しないため、対数尤度関数の最大化はニュートン法などの数値最適化手法を用いる。

新たに観測した説明変数 \mathbf{x} に対して、

$$\Pr(y=1|\mathbf{x}) = \frac{1}{1 + e^{-\hat{\mathbf{w}}^T \mathbf{x}}}$$

によって対応する応答変数が取る値が 1 である確率が得られるので、例えばこの確率が 0.5 以上の時には \mathbf{x} に対応する y は 1 であるとして、0.5 未満の場合には 0 であるとして、新しい観測が 2 つのクラスのいずれに属するかを判別することが出来る。

(3) 実験設定と結果

能動学習の設定として、初期予測モデルを fcc, bcc それぞれ 25 件のサンプルを用いて学習した後、エントロピー基準による不確実性サンプリングによって、学習データ候補の中で未使用のサンプルを一つ追加して予測器を学習しなおすことを 100 回行った。また、比較対象としてランダムにサンプルを一つ一つ追加して予測器を学習する実験も行った。図 1 は、テストデータの選択と初期モデルの学習データの選択、及びランダムにサンプルを選択する際の乱数を変化させて 20 回同じ実験を繰り返したときの、予測器の学習に用いたデータ数(横軸)と平均予測誤差の平均をプロットしたものである。実線が能動学習を用いて学習した予測器による予測誤差、点線がランダムにサンプルを選択して学習した予測器による予測誤差である。学習の初期段階(追加するサンプルが 30 件程度まで)では、能動学習によって選択されたサンプルを追加することで予測精度が劣化している。これは、少数個のデータを用いて構築した予測器ではクラス事後確率が正しく学習出来ていないため、それに基づいて選択した追加学習サンプルが予測器の性能向上に寄与していないという事情であると考えられる。一方、ある程度予測器の学習が進んだ後

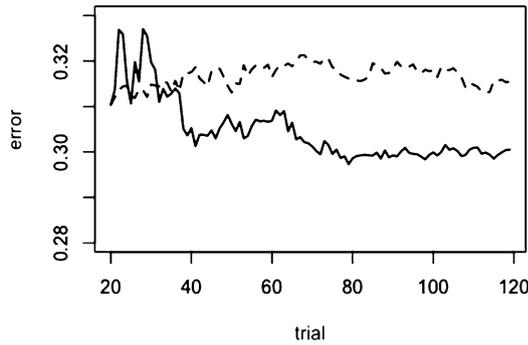


図1 能動的サンプリングに伴うテスト誤差の低減.

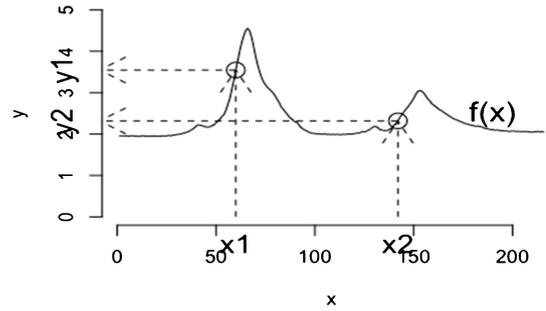


図2 多次元正規分布とガウス過程.

は、能動的に学習サンプルを選択することでランダム選択と比較して効率的に予測精度を向上させることに成功している。

(4) X線スペクトル測定の高効率化

X線スペクトル測定は試料にX線を照射し、試料から放出されるX線や電子のエネルギー分布(X線スペクトル)を調べる実験手法である。X線スペクトル測定にはさまざまな種類があるが、特にX線磁気円二色性スペクトル測定に着目する。X線磁気円二色性スペクトル測定では物質の磁気的な情報を調べることができ、永久磁石材料やスピントロニクス材料の研究に欠かせない実験手法のひとつである。近年、研究開発サイクルが加速するなかで、迅速に実験データを取得・解析し、物質・材料開発にフィードバックすることが求められている。X線磁気円二色性スペクトル測定も例外ではなく、これまでよりも測定のスループットを上げる必要がある。X線磁気円二色性スペクトル測定ではエネルギーを細かく変えながらX線を試料に照射して、対応するスペクトルを求める。X線磁気円二色性スペクトルでは「磁気光学総和則」が知られており、スペクトルの積分強度からスピン磁気モーメントと軌道磁気モーメントを定量解析することができる。従来の測定では、どの程度細かく測定するか、あるいは測定点数を何点にするかなどを予め実験者が決めていた。実験データの品質にこだわるあまり、必要以上に測定点数を多くして、貴重な実験時間を無駄に費やす可能性もある。そこで、能動学習を用いて適応的に次にどのエネルギーで測定を行うかを定め、実験者に依らず必要十分な実験データを効率的に取得する方法を考える。これによりこれまでと同等の情報を得るために必要な計測点数を削減することができて、測定時間そのものを短縮することができるようになる。ここで計測点数を合理的に減らすために「ガウス過程回帰⁽⁵⁾」を用いる。

(5) ガウス過程回帰

通常の変量ガウス分布は、平均ベクトル μ と分散共分散行列 Σ によって定まる分布であった。一方、ガウス過程はベクトルではなく関数が従う確率分布の一つであり、平均関数 $m(x)$ と共分散関数 $k(x, x')$ によって指定される。少し数学的な厳密さには欠けるが、関数が従う確率分布について

直観的な理解を目的として説明をしよう。

図2では、 xy 平面上に関数 $y=f(x)$ がプロットされている。適当な2点 x_1, x_2 に対応する $y_1=f(x_1), y_2=f(x_2)$ を考える。ここで、二次元ベクトル確率変数 (Y_1, Y_2) が平均ベクトル (y_1, y_2) 、分散共分散行列 $\Sigma_{1,2}$ で定まる正規分布に従っていると仮定しよう。このようにして、 x 軸から d 個の点を指定することで d 個の対応する y 軸上の点が定まり、それを d 次元ベクトルとみなす、そしてさらに d 次元正規分布に従う確率変数ベクトルとみなす、という操作を考える。図では2点のみであった x 軸上からのサンプルを、無限に細かくとり、最終的には x 軸のすべての値を y 軸に $f(x)$ で対応させることで、ベクトルから関数へ考察の対象を拡張することが出来る。そして、確率変数ベクトルを考えたと同様に、確率的な振る舞いをする関数を考えることも出来る。この、確率変数である関数 $f(x)$ がガウス過程であるとは、任意の個数の点 x_1, x_2, \dots, x_n における n 次元確率変数ベクトル $(f(x_1), f(x_2), \dots, f(x_n))^T$ が正規分布に従うときをいう。通常の正規分布が平均と分散によって定まると同様に、ガウス過程は平均関数 $m(x) = \mathbb{E}[f(x)]$ 及び、二点 x, x' の共分散を定義する共分散関数 $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ を指定することで定まる。このようにして確率的に振る舞う関数考えたとき、その最も基本的なものがガウス過程である。実際には観測した点や観測候補となる点を指定して得られる多次元正規分布を扱うことになるので、正規分布の基本的な性質を理解していれば問題ない。ここで必要となる正規分布の性質は、 $(\mathbf{x}, y)^T$ が平均 $(\mu_x, \mu_y)^T$ 、分散共分散行列 $\begin{pmatrix} A & \mathbf{c} \\ \mathbf{c}^T & b \end{pmatrix}$ に対して、 \mathbf{x} で条件づけた y の確率分布がやはり正規分布になり、その平均が $\mu_y + \mathbf{c}^T A^{-1}(\mathbf{x} - \mu_x)$ 、分散共分散行列が $b - \mathbf{c}^T A^{-1} \mathbf{c}$ になるという事実である⁽⁵⁾。この事実を用いて、はじめに $\{(x_i, y_i)\}_{i=1}^n$ として n 点での計測が得られている状況で新たに x_{new} という点を計測するとした場合の、 $\mathbf{f} = \mathbf{y} = (y_1, \dots, y_n)^T$ と y_{new} の同時分布は、はじめは平均関数を0としておくと

$$\begin{pmatrix} \mathbf{y} \\ y_{\text{new}} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{bmatrix} K & \mathbf{k}_{n, \text{new}} \\ \mathbf{k}_{n, \text{new}}^T & k_{\text{new}} \end{bmatrix} \right)$$

である。ここで、 $K \in \mathbb{R}^{n \times n}$ は予め観測されている n 個のデータで評価した $n \times n$ の分散共分散行列であり、 $\mathbf{k}_{n, \text{new}}$ は各

成分が $k(x_i, x_{\text{new}})$, $i=1, \dots, n$ で定義される新規データと既観測データとの共分散ベクトルであり, $k_{\text{new}}=k(x_{\text{new}}, x_{\text{new}})$ である. 同時正規分布に従う変数の条件付き分布の性質から, $\{(x_i, y_i)_{i=1}^n\}$ を観測したという条件の下で新たに x_{new} に対応する y_{new} のとる値の平均と分散はそれぞれ

$$\mathbf{k}_{n,\text{new}}^\top K^{-1} \mathbf{y}, k_{\text{new}} - \mathbf{k}_{n,\text{new}}^\top K^{-1} \mathbf{k}_{n,\text{new}}$$

となることがわかる. つまり, ガウス過程回帰を用いると, まだ観測していない点 x_{new} における y の予測値として $\mathbf{k}_{n,\text{new}}^\top K^{-1} \mathbf{y}$ が得られ, さらにその予測値のバラツキも分散 $k_{\text{new}} - \mathbf{k}_{n,\text{new}}^\top K^{-1} \mathbf{k}_{n,\text{new}}$ として定量的に評価できる.

(6) 能動的な X 線スペクトル測定

X 線スペクトル測定の目的はスペクトルという連続的なエネルギーの関数 $f(x)$ を求めることであり, この関数形状をエネルギーを x としてスペクトルを y としたガウス過程で表現(近似)することを考える. すなわち, できるだけ少数の観測点(エネルギー値)における応答 y を用いてスペクトル形状を近似するために, 能動的に応答を測定するエネルギー値を選択するという問題を考えるのである.

まず初期データとして, 少数のエネルギー点を計測する. この初期データを入力としてガウス過程回帰で「学習」すると, 出力として予測スペクトルと予測スペクトルの分散が得られる. 予測スペクトルを解析して物理量(今回は磁気モーメント)を定量評価するとともに, 分散に基づいて次に計測すべきエネルギー点を自動的に決定する. 次に計測する点の決定方法は上述のように色々と考えられるが, ここではサンプルに予測分散が最大の点を計測することとする. 正規分布の分散の対数がエントロピーと比例することから, この方法は上述の不確実性サンプリングをエントロピー基準で実行していると解釈できる. また, 分散最小化基準による選択を近似的に実行しているとも解釈できる. 物理量が設定した収束条件を満たした場合は測定を終了し, そうでない場合は次のエネルギー点を計測する. この計測データを入力に加えて再びガウス過程回帰を行う. 図3には, SmCo5 の XAS (X 線吸収分光スペクトル) を 210 エネルギー点で計測した値が点線で, 初期モデルとしてランダムな 30 エネルギー点での計測結果を用いて当てはめたガウス過程の, 未計測エネルギー点での平均が実線で, そのプラスマイナス 2 シグマが一点鎖線で示されている. 垂直の破線は, 最も予測分散が大きいエネルギー点を表しており, 次にこのエネルギー点で計測することで, 分散の大幅な低減が期待できる.

このように学習・予測・解析・計測を繰り返すことで, 今回の結果では, 従来の網羅的な計測に比べて 5 分の 1 程度の計測点数で同等の精度で物理量を決定できることが明らかになった(6).

4. おわりに

本稿では, 機械学習における教師付き学習の問題で, 例題に正解ラベルを付与するためのコストが大きい際に, より効

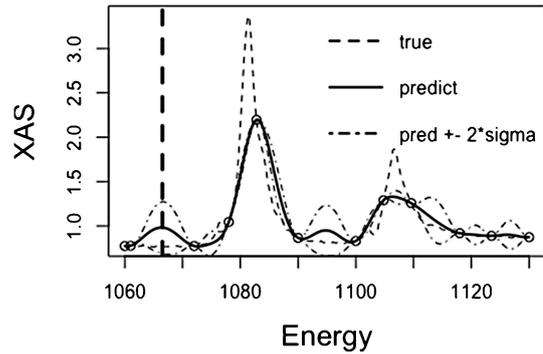


図3 ガウス過程に基づく能動的スペクトル計測.

率的に学習を行うための方法論である能動学習の基礎を説明した. 次にラベル付をするサンプルを選択するためのいくつかの基準を紹介したが, この基準で評価する対象である説明変数を完全に自由に選択できるという状況は少ないであろう. この説明変数の選択候補に関しても, 異なる問題設定がありえる. また, クエリ生成に関しても今回紹介した3種類以外にも選択基準は複数提案されている. さらに, 現実の問題では単一のクエリを生成するのではなく, 複数のサンプルを試作するほうが効率が良いこともあり, 多点探索型の能動学習も研究がされている. 能動学習は現実問題におけるラベル付コストの低減という重要な要請から, 従来から盛んに研究が行われているが, その有効性に関しては個々の問題に依存するところが大きく, 各分野でこれから知見を蓄積していく必要があると思われる.

本稿の執筆にあたり, 東京理科大学鈴木雄太氏の協力を頂きました. また, 本稿で紹介した研究の一部は, JST CREST JPMJCR1761 の補助を受けて行われました.

文献

- (1) S.D. Silvey: *Optimal Design: An Introduction to the Theory for Parameter Estimation*, Chapman Hall, (1980).
- (2) D. Cohn, L. Atlas and R. Ladner: *Machine Learning*, 15(1994), 201-221.
- (3) B. Settles: *Computer Sciences Technical Report 1648*, University of Wisconsin-Madison, (2009).
- (4) T. Hastie, R. Tibshirani and J. Friedman: *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, Second edition. Springer, (2009).
- (5) C.E. Rasmussen and C.K. Williams: *Gaussian Processes for Machine Learning*, The MIT Press, (2005).
- (6) T. Ueno, H. Hino, A. Hashimoto, Y. Takeichi, M. Sawada and K. Ono: *npj Computational Materials*, 4(2018), 4.



日野英逸

★★★★★★★★★★★★★★★★★★★★
 2005年 京都大学大学院情報学専攻修士課程修了
 2010年 早稲田大学大学院先進理工学専攻博士後期課程修了
 2013年4月-2016年3月 筑波大学システム情報系 助教
 2013年-2016年 筑波大学システム情報系 准教授
 2018年4月- 現職
 専門分野: 数理工学
 ◎数理的なアプローチによる問題解決, 特に機械学習と応用統計の研究に従事. 統計アルゴリズムの開発・理論解析と, 自然科学や産業への応用に関心を持つ.
 ★★